OPEN ACCESS

UNIVERSITY OF DUNDEE

University of Dundee

A Framework for Speechreading Acquisition Tools

Gorman, Benjamin; Flatla, David

Published in:
Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems

DOI:
10.1145/3025453.3025560

Publication date:
2017

Document Version
Final published version

Link to publication in Discovery Research Portal

Citation for published version (APA):
Gorman, B., & Flatla, D. (2017). A Framework for Speechreading Acquisition Tools. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. (pp. 519-530). Association for Computing Machinery. DOI: 10.1145/3025453.3025560

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 01. Jul. 2017

A Framework for Speechreading Acquisition Tools

Benjamin M. Gorman
University of Dundee
Dundee, Scotland
b.gorman@dundee.ac.uk

David R. Flatla
University of Dundee
Dundee, Scotland
d.flatla@dundee.ac.uk

ABSTRACT
At least 360 million people worldwide have disabling hearing loss that frequently causes difficulties in day-to-day conversations. Traditional technology (e.g., hearing aids) often fails to offer enough value, has low adoption rates, and can result in social stigma. Speechreading can dramatically improve conversational understanding, but speechreading is a skill that can be challenging to learn. To address this, we developed a novel speechreading acquisition framework that can be used to design Speechreading Acquisition Tools (SATs) – a new type of technology to improve speechreading acquisition. We interviewed seven speechreading tutors and used thematic analysis to identify and organise the key elements of our framework. We then evaluated our framework by using it to: 1) categorise every tutor-identified speechreading teaching technique, 2) critically evaluate existing conversational aids, and 3) design three new SATs. Through the use of SATs designed using our framework, the speechreading abilities of people with hearing loss around the world should be enhanced, thereby improving the conversational foundation of their day-to-day lives.

ACM Classification Keywords
K.4.2 Social Issues: Assistive technologies for persons with disabilities

Author Keywords
Speechreading; Lipreading; Hearing Loss; Deafness

INTRODUCTION
More than 11 million people (1 in 6) in the UK have some degree of hearing loss [1] and in the USA, an estimated 30 million people (12.7%) 12 years and older have hearing loss in both ears [10]. On a global scale, the World Health Organisation estimates that 360 million people (~5%) worldwide have disabling hearing loss[1] [53]. Hearing loss prevalence increases with age [10], resulting in an anticipated growth in hearing loss in the future (e.g., 1 in 5 people are expected to have hearing loss by 2035 in the UK [2]).

[1]Hearing loss greater than 40 dB in the better hearing ear in adults and greater than 30 dB in the better hearing ear in children.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

Copyright is held by the owner/author(s).
CHI 2017, May 06-11, 2017, Denver, CO, USA.
ACM 978-1-4503-4655-9/17/05.
http://dx.doi.org/10.1145/3025453.3025560

Hearing loss results in difficulties understanding what others are saying during conversation [49]. Considering that the quality of our communication, our willingness to engage in conversation, and our level of apprehension about conversations all profoundly affect the quality of our lives [18], conversational difficulties due to hearing loss can result in social isolation [22], career stagnation [39], and a decrease in overall life satisfaction [49]. Hearing aids are intended to help reduce conversational difficulties, but are expensive [11], often counterproductive in noisy environments [25], and have low adoption rates (~14% [1]) due to a lack of comfort [40] and perceived social stigma [30].

People who are deaf or have a hearing loss find that speechreading (commonly called lipreading) can overcome many of the barriers when communicating with others [18]. Speechreading can be described as a special case of audio-visual speech recognition where emphasis is placed on the visible, rather than on the audible, speech information. Speechreading helps to improve conversational confidence (thereby reducing social isolation), enhance employability, and improve educational outcomes [49]. Speechreading also does not rely on the other conversation partner's knowledge of a Signed Language or a technique such as Cued Speech [19]. Speechreading is also partially used by those without hearing loss (albeit subconsciously) [26], so it is already a natural component of conversation, thereby limiting the chances for stigmatisation.

However, speechreading is a skill that takes considerable practice and training to acquire [34]. Publicly-funded speechreading classes are sometimes provided, and have been shown to improve speechreading acquisition [5]. However, classes are only provided in a handful of countries around the world; there is an insufficient number of classes running in areas in which they are provided (e.g., only 50 of an estimated 325 required classes are currently running in Scotland [5]) and classes require mobility to attend.

To help expand speechreading training worldwide, we developed a novel framework that can be used to develop Speechreading Acquisition Tools (SATs) – a new type of technology designed specifically to improve speechreading acquisition. Through the development and release of SATs, people with hearing loss will be able to augment their class-based learning, or learn on their own if no suitable classes are available.

To develop our framework, we conducted in-depth interviews with 7 of the 21 Scottish speechreading tutors [5] to explore their background, approach to teaching, current use of technology, and thoughts on how speechreading can be improved.

Through thematic analysis of our interview transcripts, we identified four main themes relevant to the future development of speechreading acquisition tools: 1) speechreading as a skill, 2) access to speechreading, 3) teaching practices, and 4) attitudes to technology. Using these themes, we developed our framework, which consists of two dimensions (*Type of Skill* and *Amount of Information*), each with three levels (*Analytic*/*Synthetic*/*Hybrid* and *Low*/*Medium*/*High*, respectively).

To evaluate our framework, we used it to: 1) classify every existing speechreading teaching technique identified by our participants, 2) critically reflect on previously-developed solutions, and 3) design three new SATs for enhancing speechreading acquisition and proficiency.

By employing our framework in this fashion, we show that it: 1) comprehensively reflects existing speechreading teaching practice, 2) can be used to help understand the strengths and weaknesses of previously-developed solutions, and 3) can be used to identify clear opportunities for the development of new SATs to help improve speechreading skill acquisition. Through the dissemination and adoption of our framework into the research community and assistive technology commercial sector, we foresee new technology being developed to help improve speechreading acquisition worldwide. Once in the hands of people with hearing loss, SATs will help enhance their speechreading capabilities, increasing their conversational confidence and reducing their social isolation.

This paper makes five contributions: First, we contribute novel speechreading tutor interview data and thematic analysis of that data. Second, we describe our speechreading acquisition framework for developing new technological tools for improving speechreading acquisition. Third, we evaluate our framework by fitting existing teaching techniques and existing solutions into the framework. Fourth, we define and describe Speechreading Acquisition Technologies (SATs). Fifth, we outline how to use our framework by designing three new SATs to improve speechreading acquisition.

BACKGROUND & RELATED WORK

Speechreading

The atomic unit of spoken language is the *phoneme*. Phonemes are combined to form words. For example, */b/*, */æ/*, and */t/* are the phonemes for "bat". There are 48 commonly-recognised phonemes in the English language [46]. For each phoneme, a speaker's lips, teeth, and tongue produce a visual representation known as a *viseme* [23].

People with hearing loss try to map visemes to phonemes in order to understand what a speaker is saying [9]. Hearing loss causes the audible phoneme to be lost or difficult to perceive, but the viseme is still available. For example, */l/* and */r/* are acoustically similar in English (especially when following another consonant, such as 'grass' vs. 'glass'), but are generated using distinct visemes, so this visual difference can be used to determine if a speaker has said */l/* or */r/*.

Although 'reading' visemes to understand what someone is saying works for many phonemes, there are also phonemes that are represented by similar visemes [36]. For example, */v/* is a voiced phoneme (the vocal cords vibrate), which is audibly distinct from */f/*, which is not voiced. However, the viseme for */v/* is very similar to the viseme for */f/*, making the words *'fan'* and *'van'* difficult to distinguish visually.[2] As a result, novice speechreaders (e.g., someone who has received no formal speechreading training) often find it difficult to fully understand what a speaker is saying, resulting in confusion, frustration, and reduced conversational confidence [18].

Conversation Aids & SATs

We define a *Conversation Aid* as any technique or technology that enables or supports face-to-face conversation. We describe existing conversation aids below in the second evaluation of our framework, however at this stage it is important to recognise that *Conversation Aids* are a superset of *Speechreading Acquisition Tools*, as the latter have been designed specifically to be used by speechreaders.

Speechreading Teaching Methods

Early speechreading training was focussed on skill acquisition as a by-product of learning speech production [24]. The explicit teaching of speechreading in classes with methodologies that were primarily concerned with speech recognition did not occur until the 19th century [32].

Speechreading classes primarily focus on learning how different mouth shapes are produced during speech [34], as well as how to use conversational repair strategies to gain important contextual information to help 'fill in' any gaps in understanding [34]. Classes also include information about hearing aids or other assistive listening devices, and give people a social space to meet with others who have a hearing loss [49]. Classes can also improve an individual's self-confidence [8], and help attendees become more knowledgeable about their hearing loss and how they can make communication easier.

There are two main approaches to teaching speechreading: *synthetic* and *analytic*. Synthetic methods use a 'top-down' approach in which the speechreader is encouraged to focus on the gist of the conversation to help them determine individual words. Synthetic methods are often referred to as 'mind-training' or 'context-training', as they rely on lexical ability to understand the topic of the conversation. Synthetic methods consider the sentence to be the basic unit of speech [49].

Analytic methods use a 'bottom-up' approach in which the speechreader is encouraged to focus on individual speech movements to identify the word, phrase or sentence [49]. Analytic methods are often referred to as 'eye-training' as the speechreader focusses on the visual aspects of a speaker to disambiguate visual speech patterns. Analytic methods hold that the syllable or phoneme is the basic unit of speech [32].

The Nitchie Method [41] was initially developed using an analytic approach, however it shifted towards a synthetic approach in later years. Nitchie is credited with establishing the foundations of modern speechreading training as well as the synthetic approach to speechreading. His method stresses that eye-training materials and those based around association or

[2] We encourage readers to make *'fffffff'* and *'vvvvvv'* sounds to hear the difference, and to make these sounds plus 'fan' and 'van' in front of a mirror to see the lack of visual difference.

ID	Gender	Age	Hearing Loss	Cause	HA/CI Use	Years Signing	Years Teaching
P1	F	78	Profound Hearing Loss	Childhood Illness	Cochlear Implant	32	32
P2	F	68	Severe Hearing Loss	Childhood Illness	One Hearing Aid	None	14
P3	F	61	Typical Hearing	-	None	None	12
P4	F	57	Typical Hearing	-	None	None	18
P5	F	42	Typical Hearing*	-	None	3	6 Months
P6	F	67	Moderate Hearing Loss	Not Specified	One Hearing Aid	None	7
P7	F	66	Severe Hearing Loss	Congenital Hearing Loss	Two Hearing Aids	19	1

Table 1. Summary of participant demographics. HA/CI = Hearing Aid/Cochlear Implant. Years Signing = Number of Years Using Sign Language (BSL). Years Teaching = Number of Years Teaching Speechreading. '*' indicates severe hearing loss for eight months due to viral infection.

context be separated. He also believed that context training was more important than training for visual disambiguation.

The Kinzie Method [35] uses a synthetic approach to speechreading and includes mirror practice (where the student talks before a mirror to learn visual differences between speech movements) and the use of voice. Materials in this method are organised into graded lessons for both children and adults, with sentences forming the basis of instruction [35].

The Mueller-Walle Method [15] is an analytic approach that emphasises eye-training through syllable drills, which are rhythmic drills consisting of contrasting syllables that are restricted to sound combinations found in the English language. Syllable drills are spoken as quickly as possible by the tutor.

The Jena Method [17] is an analytic approach that emphasises syllable drills and stresses kinaesthetic awareness during speech production as well as audition and eye-training. Eye-training is accomplished through syllable drills in a similar manner as the Mueller-Walle and early Nitchie Methods. During the drills, the speaker is expected to speak in unison with the instructor and imitate their lip and jaw movements thereby concentrating on the kinaesthetic sensations experienced.

Although the methods described above give us a theoretical understanding of the approaches to speechreading training, we wanted to base our framework on current practice (which is most likely influenced by these theories). As such, we conducted interviews with practicing speechreading tutors to generate the dataset needed to develop our framework.

METHOD

We conducted individual interviews with seven Scottish speechreading tutors [5] to explore their background, approach to teaching, current use of technology, and thoughts on how speechreading can be improved. With participant's informed consent, we audio recorded each interview, then transcribed and thematically analysed [13] the transcripts.

Aims

We had five main aims (phrased as questions) guiding our interviews: 1) Do speechreading tutors primarily employ analytic or synthetic training approaches? 2) What do speechreading tutors consider to be the basic unit of speechreading? 3) What technology (if any) do speechreading tutors currently use to teach speechreading? 4) Do speechreading tutors feel that speechreading training could be improved with new technology or training techniques? 5) How do speechreading students continue to learn when not in class? Our interview guide is included in the supplementary materials ('InterviewGuide.pdf').

Participants

Scotland is the one of the few countries that provides formal training for speechreading tutors. As such, our participants all reside in Scotland, and offer classes throughout central Scotland. Participants were recruited through direct emails via contact details obtained from the Scottish Course to Train Tutors of Lipreading (SCTTL, http://www.scotlipreading.org.uk/), along with word-of-mouth. Participant details are summarised in Table 1. All participants were female (aged from 42 to 78). The 21 tutors listed on SCTTL were all female, therefore we could not balance genders. The range of teaching experience was from six months to 32 years. Our participants self-reported details about their hearing, which we classified into five levels using the textual descriptions of hearing loss identified by Action On Hearing Loss [3]: Typical Hearing, and Mild, Moderate, Severe, and Profound Hearing Loss. The participants' backgrounds were varied:

P1, 78, has had profound hearing loss from a young age. She has been teaching speechreading for 32 years after being asked to take over a class by a friend. She has no formal teacher or speechreading training. P1 teaches three classes a week.

P2, 68, has severe hearing loss. She has been teaching speechreading for 12 years. She was a high school science teacher who retired early due to her hearing loss. She had attended speechreading classes for around two years before her tutor convinced her to take the SCTTL so that she could teach classes herself. P2 teaches two classes a week.

P3, 61, has typical hearing. She has been teaching speechreading for 12 years. She previously worked as a subtitler for the BBC. Her father had a hearing loss which motivated her to become a speechreading tutor. She undertook the SCTTL at the same time as P2. P3 teaches two classes a week.

P4, 57, has typical hearing. She has been teaching speechreading for 18 years. She also works as a Speech and Language Therapist for the Scottish National Health Service. P4 teaches one class per week and also trains tutors as part of the SCTTL.

P5, 42, has typical hearing, however she experienced severe hearing loss for eight months after a viral infection. She has been teaching speechreading for six months. Her motivation for teaching started after experiencing hearing loss. P5 initially wanted to learn to sign to increase access for deaf individuals at her community centre where she works. However, upon

realising that the deaf community in her area instead required a speechreading tutor she undertook the SCTTL. P5 teaches two classes a week.

P6, 67, has moderate hearing loss. She has been teaching speechreading for seven years. She temporarily lost her hearing due to a viral infection 10 years ago, so her daughter encouraged her to learn speechreading to help her cope. After a year within the class, her tutor encouraged her to take the SCTTL. P6 teaches four classes a week.

P7, 66, has severe hearing loss. She has been teaching speechreading for one year. Her motivation for becoming a speechreading tutor was to help individuals who experience hearing loss later in life; she worked in a library and noticed that these individuals seemed prone to isolation. She also took the SCTTL. P7 teaches three classes a week.

Approach

After obtaining ethical approval, we conducted semi-structured one-to-one interviews. Interviews took place in mutually-convenient and quiet locations. The mean interview time was 40 minutes (max 74 minutes, min 28 minutes); some interviews took longer due to participants discussing past students and their experiences interacting with them. The interviews consisted of questions that explored their background, general teaching approach, assessment, current use of technology and their thoughts on where speechreading could be improved in the future. The interview questions were open-ended (for example, "Why did you decide to become a lipreading tutor?" [3]) and we let the participant lead whenever possible [7], encouraging elaboration by asking probing follow-up questions when necessary [7] (e.g., "And do they find that helpful?"). Audio recordings were gathered during the interviews. Transcripts were coded and analysed using thematic analysis [13], grouping similar experiences together in order to identify themes across all participant interviews. The themes were refined through an iterative thematic analysis process to generate a final, distinct set of themes.

Phases Of Analysis

Phase 1: Becoming Familiar With the Data

We manually transcribed the interviews using custom-built software. We strove for a verbatim account of all verbal and nonverbal (e.g., laughs) utterances. We added punctuation where necessary to indicate pauses, full sentences, and questions. We formatted our transcripts using 'Interviewer:' to indicate interview statements and 'Participant1:' to indicate statements by P1. All names and locations were removed to maintain anonymity. Our interview transcripts are included in the 'Transcripts' folder in the supplementary materials.

Phase 2: Generating Initial Codes

We started by reading paper copies of the transcripts and manually highlighting all interesting extracts. We then transferred our highlighted extracts into MAXQDA12 [4], resulting in 502 extracted elements from the original transcripts.

Using the 'coded segment' feature of MAXQDA12, we then systematically processed our original 502 elements iteratively using a data-driven approach to ensure that our final themes emerged exclusively from our interview data.

During this process, we coded for maximum diversity of potential themes and patterns, and did not discard data unless it was clearly not relevant to the research (e.g., an unrelated anecdote). We also kept enough of the text around each coded segment to retain the segment's context, because a common criticism of coding is that the context is often lost [16].

By giving equal attention to each extract, we further segmented our extracts and coded them using iteratively-shaped codes, resulting in 944 coded segments under 116 unique codes. Our MAXQDA12 data file containing our final coded segments and codes is included in the 'Analysis.mx12' file in the supplementary materials.

As we followed Braun and Clarke's [13] 15-point checklist of criteria for good thematic analysis, we did not conduct inter-rater coding. Inter-rater reliability checks are not always used in thematic analysis since there is scepticism regarding such tests [4, 42, 50]; it can be argued that one researcher merely trains another to think as she or he does when looking at a fragment of text and so the reliability check does not establish that the codes are objective but merely that two people can apply the same subjective perspective to the text [37].

Phase 3: Searching for Themes

In this phase, we created a short definition for each code that described when that code would be used and what it represented across the entire data set. We then printed each code plus its definition on individual strips of paper, and iteratively organised the strips into 'theme-piles' on a whiteboard. Using the resulting 'theme-piles', we produced an initial thematic map. Our initial thematic map is in the 'InitialThematicMap.png' file in the supplementary materials.

Phase 4: Reviewing Themes

Starting with our initial thematic map from Phase 3, we removed themes that did not directly relate to the study aims outlined above. We then reviewed the collection of coded extracts for each remaining theme to ensure that they formed a coherent pattern. Finally, we re-read our original transcripts to check that our revised themes provided suitable coverage.

Phase 5: Defining and Naming Themes

Once our thematic map was finalised (Figure 1), we defined each theme by examining its collection of coded extracts to determine the main aspect of the data captured by the theme. We then revisited the collection of coded extracts for each theme, refining the story told by each theme. Finally, we produced internally-consistent accounts, with an accompanying narrative for each theme.

FINDINGS

Through our thematic analysis, we identified four themes: 1) speechreading as a skill, 2) access to speechreading, 3) teaching practices, and 4) attitudes to technology. We now explore each theme in detail using quotes from participants to scaffold the narrative of each theme.

[3] In the UK, 'speechreading' is referred to as 'lipreading', therefore in discussions with participants we used the term 'lipreading'.

[4] A qualitative analysis software package, http://www.maxqda.com/

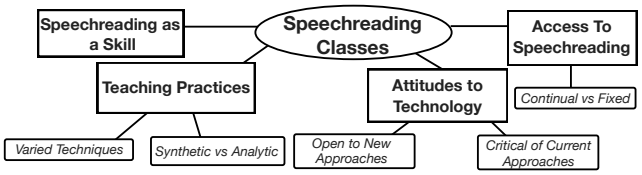

Figure 1. Final thematic map of four main themes and their subthemes.

Speechreading as a Skill

Our participants saw speechreading as a skill, one that requires long periods of concentration and focus to learn. Classes are typically two hours long with a short break, so students focus for around an hour at a time. All seven participants discussed the need for concentration and focus within classes:

> P4: *"For the person themselves they need that confidence, that assertiveness, to do that, they also need to concentrate and pay attention for that length of time."*

> P6: *"...not looking around the room and trying to listen but actually just focusing...the amount of concentration these people give, it's amazing...it's as almost as they are drilling holes in you...it's excellent, it's really good."*

As illustrated by these comments, a student's ability to focus and concentrate is of paramount importance to learning speechreading. A high level of concentration is necessary due to the limitations of speechreading – many aspects of speech (e.g., voicing) are mostly audible instead of visible:

> P2: *"I warn them beforehand, that only 1/3 of speech is lip-readable...they are aware that there is a limitation to what we are doing but it's an added help to everything they do...it's useful but it doesn't solve all your problems."*

In addition to the level of concentration required and the limitations of speechreading, participants also described additional factors that pose difficulties for students, such as particular words or speech movements having little visual difference:

> P6: *"Knowing that...some of these skills are very subtle, you are not going to see huge [differences]. Some speech movements are very clear, [but] when you get to others there are some sounds that are so subtle you can hardly see them...vowel at the beginning of the word, like 'ahead'...it's difficult to spot. Sometimes if you can't lipread, [it] could be that there is a vowel in there that you are not aware of."*

> P1: *"Certain words and sounds they find really difficult. I have to repeat them...I go over things quite a lot."*

Many participants identified how different accents can affect individual speech movements as well:

> P4: *"Sometimes you've got...English [as opposed to Scottish] vowel sounds coming at you and therefore sometimes you look and think 'I don't know what that is'. Because it's got the accent it kind of changes things."*

The challenges of speechreading, plus the level of concentration required, often lead to fatigue during and after class:

> P2: *"And there are times when you get tired and your lipreading is absolutely rubbish."*

> P4: *"As to how tiring it is to sit and watch somebody for...two hours, we build in breaks but it's still a lot of effort and concentration. It's very tiring and I think that's what comes after two or three weeks...somebody will come up to you and say I went home and I was absolutely exhausted. I didn't realise how tiring it was."*

Access to Speechreading

Access to speechreading classes was discussed by all participants. In particular, participants focussed on issues surrounding funding of classes, the length of classes and how students beginning classes (and the general public) have little awareness of what is involved in learning speechreading:

> P4: *"...it's a very difficult one, lipreading is...kinda like a Cinderella Service [ignored or treated as less important]. People don't recognise that actually everybody lipreads to a certain extent. I think what could be improved with lipreading is general awareness of the fact that everybody lipreads, so...more people would be aware of it and more people would therefore come to the classes."*

Participants also highlighted that local governments can view speechreading classes as recreational, causing funding issues:

> P7: *"I would prefer it to come under university rather than sitting under [local governments], because it is a life changing skill...rather than a hobby or a job."*

> P2: *"Some authorities regard it as a recreational thing"*

Continual vs. Fixed

A subtheme within access reflects whether speechreading classes should be continual or fixed length. All tutors agreed that classes should be continual.

> P1: *"I do think that judging by this and my [other] class that it's important for the class to be ongoing unless they for individual reasons want to drop out."*

> P2: *"It's a continual practice thing...you could really do with a little practice every week...for the rest of your life."*

However, many of the tutors teach fixed-length classes. This was generally due to funding issues, with many local governments only offering two years of speechreading classes:

> P6: *"I know, two years is our maximum and then you have to go on to a paid class."*

> P2: *"So I'm paid by [anonymised] city council. They provide two years free lipreading, I am the only tutor in [anonymised] that does it. I did one on the Tuesday, two on a Thursday, but the budget is decreasing every year. We don't know how long this will go on for. I only teach for thirty weeks...that's all the council will pay you for."*

> P5: *"This is a problem for me especially because I have got funding from [anonymised], and they will not perpetually fund something. So I have got funding for one year of lipreading classes, which is a 30 week course."*

In some cases, local governments offer no funding for classes, and students living in these areas pay for classes.

> P3: *"They do pay in the [anonymised] groups. In the other groups no they didn't pay.*

Students also appear to be willing to attend continual classes rather than fixed length classes. This is supported by participants reporting that students only stop attending classes due to becoming ill or passing away.

> Interviewer: *"So why do people stop coming to a class?"*
> P7: *"I haven't had any experience of that, they have all been very faithful including [over] the holidays."*

> P1: *"Some people do drop away. Usually they either die or [grow] too old. Some people died in this class and they died in my [anonymised] class too. Not many have just dropped out."*

> P6: *"People who come tend to stay. A gentleman stopped coming to this class about a year ago. He is...very old, in his 90's. He lives about 10 miles down the road and has to come by bus and therefore it's very difficult for him."*

Teaching Practices

Teaching practices varied widely, with all tutors using a variety of approaches and techniques to teach speechreading.

Synthetic vs. Analytic

Synthetic and analytic are the two main approaches to teach speechreading. Although the teaching methods described in our Background section typically emphasise one over the other, all of our tutors draw from both approaches as needed:

> P4: *"We tend to have a general topic for the class...if I am doing something on a movement, 'f'...I might choose a topic that begins with that sound...I might talk about 'fireworks' or 'fire'. The whole lesson will not be around that particular speech sound...I would do a follow up story or exercise with that sound appearing it in regularly."*

> P6: *"If you know the context, you can make a really good guess, which is a lot of lipreading anyway. If they are talking about horse racing then there is not going to be anybody talking about ballet dancers, it's unlikely. Knowing that it's very subtle, that some of these skills are very subtle...some speech movements are very clear, when you get to others there are some sounds that are so subtle you can hardly see them."*

Varied Techniques

Tutors did not mention the teaching methods introduced earlier, but discussed techniques that can be linked to these methods. Participant descriptions of techniques were often fragmented throughout an interview, resulting in many relevant quotes. We provide condensed summaries (without using quotes) of these descriptions below and refer readers to our supplementary materials for more details.

Mirror Practice: student speaks into a mirror to see his/her own mouth shapes and the visual differences between speech sounds. Used by all seven tutors. Analytic.

Quick Recognition Exercises (QREs) or *Syllable Drills*: rhythmic drills consisting of contrasting syllables or words spoken as quickly as possible in different orders by tutor. Class members repeat back. Used by all seven tutors. Analytic.

Cue Recognition: encompasses looking for body language, facial emotion, and hand gestures. Used by all seven tutors. Analytic and synthetic – Hybrid.

Speech Movements or *Lip Shapes*: focussed attention on the visual representation of a single speech sound on its own or within a word. Used by all seven tutors. Analytic.

Pair Work: focussed speechreading practice in pairs. Used by all seven tutors. Hybrid.

Stories: tutor speaks a multi-sentence story with known topic to the class. Used by P2, P3, P4, P5, P6 and P7. Synthetic.

Fingerspelling: spelling spoken words using sign language hand movements. Tutors typically sign only the first speech sound of a word. Used by P2, P4, P5 and P7. Analytic.

Scenarios: basing a lesson around pretending you are in a specific place such as the dentist, so all material is based on that scenario. Used by P1, P2, P4 and P5. Synthetic.

Word Quizzes: based around a topic, such as 'animals', and students having to watch for an animal for each letter of the alphabet. Used by P2, P5 and P7. Hybrid.

Framed Sentences: tutor says a topic-based sentence such as *'In my garden I have ____'*, but mouths the blanked word. Used by P2 and P5. Synthetic.

Mystery Object: an unknown object is hidden using paper or cloth. The tutor speaks a number of 'hint' sentences to aid identification. Used by P2 and P3. Synthetic.

Kinaesthetic awareness was the only technique tutors discussed that was not used to teach speechreading. P4 and P7 mentioned that it was useful when teaching students aspects of speech production such as controlling the loudness of their voice, but did not mention its place within speechreading.

Attitudes To Technology

Participants appear to use some non-specialised technology in their classes (e.g., hearing aids, loop systems, videos, PowerPoint) and five tutors reported informing students of Lipreadingpractice.co.uk, a website containing videos of some of the exercises used within classes. It is interesting to note here that recently-trained tutors were more critical of existing practices and were more open to using new technology.

Critical Of Current Approaches

Participants were critical of current approaches to learning speechreading using technology. Subtitles were praised for allowing individuals to enjoy videos, however tutors also reported that subtitles do not improve speechreading as students have to either watch the subtitles or watch the video:

> P4: *"I never advocate watching television to practice lipreading. If you have ever tried...it's horrendous!"*

Lipreadingpractice.co.uk was also criticised for having distracting videos and limited training material:

P2: *"The trouble with a lipreading site like that is there are only so many stories. Eventually, if...you've got a good memory, you are just gonna know them all."*

Open To New Approaches

Participants were open to new approaches involving technology and several discussed how mobile apps could be developed to help students practice outside of classes:

Interviewer: *"Do you think that people who are looking into learning lipreading would benefit from additional kinds of new technology to help?"*
P3: *"I would think so yes. But I don't have the technological ability so I can't tell you what it would be. But I'm sure because there is so much out there, iPhones, iPads and you know tablets and all that. Surely to goodness there must be something that we can do that will help."*

P5: *"I mean I don't know if there [are] lipreading apps. I think a lipreading app would be good. Just for people to practice...apps are the way forward aren't they? How you would develop it I'm not quite sure...it would just need to be about the shapes and practicing the words and having a bit of context..."*

FRAMEWORK

Our themes highlight how speechreading acquisition can be enhanced through the development of new assistive tools that will help resolve issues regarding the lack of access, and can also be specialised to different speechreading teaching approaches. We call these tools Speechreading Acquisition Tools (SATs) – a new type of technology designed specifically to improve speechreading acquisition. We believe that through the development and release of SATs, people with hearing loss will be able to augment their class-based learning, or learn on their own if they are unable to attend a speechreading classes.

To facilitate the transition of knowledge from our thematic analysis to researchers, Assistive Technology (AT) commercial sector, and AT enthusiasts, we developed a *speechreading acquisition framework*. Our framework can be used to design SATs by individuals who may have a varied amount of knowledge of speechreading, phonetics and hearing loss.

Framework Design

Our thematic analysis identified numerous speechreading teaching techniques with each technique emphasising analytic, synthetic, or hybrid (analytic and synthetic) speechreading skills. Each teaching technique also provides a different amount of information, so that they can be chosen to complement a student's current speechreading ability.

Identifying that 'skill type' and 'information amount' apply to each teaching technique, we used these as the base dimensions for our framework: *Type of Skill* and *Amount of Information*. The dimensions are continuous in nature, but to improve framework accessibility we discretised each into three levels (*Analytic*/*Hybrid*/*Synthetic* and *Low*/*Medium*/*High*, respectively), resulting in a 3x3 cell-based grid as shown in Figure 2.

We classify a technique as *Analytic* if it focusses on visual disambiguation. A technique that focusses on leveraging the context is classified as *Synthetic*. *Hybrid* techniques focus on both analytic and synthetic approaches to speechreading.

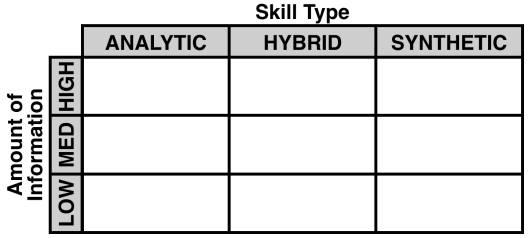

Figure 2. Our speechreading acquisition framework, with two dimensions: *Type of Skill* and *Amount of Information*, each split into three levels (*Analytic*/*Hybrid*/*Synthetic* and *Low*/*Medium*/*High*, respectively)

As the basic unit of analytic teaching methods is the phoneme, analytic techniques that provide individual phonemes are classified as *Medium*, techniques that provide non-phonemic information (e.g., speech production properties) are *Low*, and techniques that provide more than phonemes (e.g., whole words) are *High*. The basic unit of synthetic teaching methods is the sentence, so synthetic techniques that provide the topic of a specific sentence are classified as *Medium*, techniques that provide less (e.g., the topic of a conversation) are *Low*, and techniques that provide more information (e.g., the topic and context of a sentence) are *High*.

FRAMEWORK EVALUATION

To evaluate our framework, we now use it to 1) classify existing teaching techniques, 2) critically reflect on previously-developed solutions, and 3) identify and describe three new technologies for enhancing speechreading acquisition and proficiency. By employing our framework in this fashion, we show that it 1) comprehensively reflects existing speechreading teaching practice, 2) can be used to help understand the strengths and weaknesses of previously-developed solutions, and 3) can be used to identify clear opportunities for the development of new SATs to improve speechreading acquisition.

Teaching Techniques

To evaluate our framework, we first fit existing teaching techniques within our framework cells. The goal of this evaluation is to assess our framework's coverage; accommodating every teaching technique identified by our interview participants indicates good coverage, any teaching techniques not fitting within our framework indicate framework incompleteness.

Fitting Identified Teaching Techniques

We now describe our classification rationale for each teaching technique described in our Teaching Practices section above. Our classifications can be seen in Figure 3.

Speech Movements and *Lip Shapes*: **analytic** and **low information** as the student is only informed of the target speech-movement(s) and has to identify the rest of the word(s).

Finger Spelling: **analytic** and **medium information** as the first letter is signed but the rest of the word has to be identified.

QREs: **analytic** and **high information** as the student knows the words or syllables spoken, so only has to order them.

Amount of Information	ANALYTIC	HYBRID	SYNTHETIC
HIGH	Mirror Practice, QREs	Word Quizzes (name the animals)	Framed Sentence Exercises
MED	Finger Spelling (initial letter)	Cue Recognition (body language)	Scenarios
LOW	Speech Movements, Lip Shapes	Pair Work	Mystery Object*, Stories

Figure 3. Placement of teaching techniques into our framework. The '*' indicates the starting amount of information provided by this technique (Mystery Object), but this level increases as more clues are given.

Mirror Practice: **analytic** and **high information** as the student knows the words and movements they are speaking.

Pair Work: **hybrid** and **low information** as the student may not be familiar with their partner's facial movements, but may know the conversation topic.

Cue Recognition: **hybrid** and **medium information** as facial expressions and hand gestures provide a substantial amount of context to an utterance.

Word Quizzes: **hybrid** and **high information** as the student is given the first letter of a word within a precise topic.

Stories: **synthetic** and **low information** as only the topic of the story is given.

Mystery Object: **synthetic** and **low information** as the student is only aware of the shape of the object hidden beneath the cover, however information increases as more clues are given.

Scenarios: **synthetic** and **medium information** as a specific context is provided and all material is based on that context.

Framed Sentence Exercises: **synthetic** and **high information** as the student is told the sentence and guesses one word; the sentence provides a high degree of contextual information.

As shown in Figure 3, our framework accommodates all of the identified teaching techniques with no gaps, thereby increasing our confidence in the coverage of our framework.

Existing Conversation Aids & SATs

In the second stage of our evaluation, we fit existing conversation aids and SATs into our framework to get a deeper sense of the coverage of our framework. Through this process we also critically reflect on the design of existing conversation aids and SATs as well as identify where new opportunities lie.

Fitting Existing Conversation Aids & SATs

We now briefly describe existing conversational aids and SATs identified via literature search, and provide our reasoning for where we fit each in our framework (placements in Figure 4).

Signed Languages (e.g., American and British Sign Language) are languages that use hand, arm, and facial gestures to facilitate communication. A signed language provides **high information** and is a **hybrid** approach as it relies on analytic skills to understand unfamiliar names and words (typically communicated using finger spelling) and synthetic skills (e.g., facial expressions) to understand particular aspects of a conversation (e.g., identifying a question). *Cued Speech* [19] is a system of eight hand-shapes placed in four positions around the mouth that aim to clarify lip-patterns during speechreading. As such Cued Speech is **analytic** and provides **high information** as its cues disambiguate all phonemes.

In spite of their benefits and sign language's importance to Deaf culture [44], sign language (cued speech) needs to be known by both conversation parters in order to help; it does not help conversations with people who do not know the language.

The earliest example of technology aiding speechreading can be seen in *Upton's Eyeglasses* [51]. This SAT used a clip-on microphone to detect speech, and processed the signal via high- and low-pass filters to classify spoken phoneme components. An LED matrix was positioned at the side of a pair of modified eyeglasses, and its light output was channeled so that it appeared at the centre of that side's lens, enabling an early augmented reality system (e.g., the bottom LED illuminated when a phoneme was voiced, making it appear as if the speaker's throat was glowing). Due to their focus on speech components, Upton's Eyeglasses are **analytic** and provide **low information**, as the wearer is only provided information about how a sound is produced. Much later, a similar peripheral-display approach was taken by *Ebrahimi* et al. [21].

Similar to Upton's Eyeglasses, *Tactile SATs* provide spoken phoneme information using tactile feedback. One such display of voice fundamental frequency showed a 10% improvement in a speech discrimination task [12], however a later study found positive results in terms of identifying voicing and for consonant identification, but no benefit for speechreading words in sentences [55]. As they are very similar to Upton's SAT, *Tactile SATs* are **analytic** and provide **low information**.

iBaldi [38] is an iOS application that overlays a visualisation of speech components onto an animated talking head. The visualisation shows one of three coloured discs (nasality in red, friction in white, voicing in blue) by iBaldi's mouth when he makes the corresponding sound. *iBaldi* provides similar information as Upton, so it is **analytic** and **low information**.

SATs that focus on helping identify components of speech based on how they are produced (Upton, Ebrahimi, Tactile, iBaldi) contradict typical speechreading approaches by training the speechreader to focus on *auditory* aspects of speech, rather than *visual*. Even though these aids can provide rich information, this information is of limited value to someone whose understanding of speech is primarily visual, not audible.

Spectrograms visualise frequency (Y-axis) over time (X-axis), with intensity mapped to colour. Linguists can use them to identify words, but this requires extensive training [28]. *Watanabe* et al. [52] improved spectrograms by integrating different speech features into a single image, but the evaluation used participants with extensive spectrogram-reading experience, so the technique's generalisability to speechreading is unknown. Both examples of *Spectrograms* are **analytic** aids, and provide **high information** due to the richness of the visualised data.

Lip Assistant [54] is an SAT that generates magnified realistic animations of a speaker's lips that are superimposed on the bottom left of a video. *Lip Assistant* is **analytic** as it focusses exclusively on lip shapes and it provides **low information**.

Amount of Information	ANALYTIC	HYBRID	SYNTHETIC
HIGH	Cued Speech, Spectrograms	Subtitles, Signed Languages	
MED		LipreadingPractice	CME (C)
LOW	Upton, Ebrahimi, TactileSATs, CME(A) Lip Assistant, iBaldi		CME (B) DAVID

Figure 4. Placement of conversation aids/SATs into our framework.

Subtitles (captions) present the speech (analytic) and sound effects (synthetic) of video as on-screen text. As such, they are a **hybrid** approach and provide **high information**, however they also require the viewer to split their attention between reading the subtitles and watching the video content; one eye-tracking study found that participants spent ~84% of their viewing time focussed exclusively on subtitles [33]. The design of Lip Assistant also suggests it is sensitive to split attention.

ConversationMadeEasy [48] is an SAT comprised of three programs, each presenting videos of speakers with or without audio. The programs increase in complexity: Program A is for **analytic** training with **low information**, and program B is for **synthetic** sentence training providing **low information**. Program C is for **synthetic** scenario-based training with commands or questions based on the scenarios given within a closed response set of four pictured options. As such, Program C provides **medium information**.

DAVID [47] is an SAT offering videos of sentences on everyday topics, such as 'going shopping'. The student watches and responds by typing the complete sentence or content words, or via multiple choice. DAVID also provides repair strategies such as repeating the sentence, or presenting words in isolation. DAVID is a **synthetic** SAT providing **low information**.

Lipreadingpractice.co.uk is a website-based SAT offering subtitled videos of consonants, vowels, and passages. The speaker says these with and without voice, shown from the front and from a profile angle, and repeats each a number of times. Words and phrases are provided as written exercises. *Lipreading Practice* is a **hybrid** SAT providing **low information**.

As shown in Figure 4, our framework can accommodate existing conversation aids and SATs, increasing our confidence in its coverage. More importantly, we also located gaps in the previous work. We next use our framework to design three new SATs that address these gaps, but also reflect the knowledge gained via our thematic analysis presented earlier.

Framework-Inspired SAT Examples

To further evaluate our framework we now demonstrate how it can be used to design new SATs. Below, we explain how we used our framework to design and prototype one speechreading acquisition tool, as well as to propose two additional SATs that we plan to build and evaluate as future work.

PhonemeViz

There are currently no **analytic** SATs that provide **medium information** to speechreaders during conversations. In addition, no SATs map to how fingerspelling is used within classes.

PhonemeViz places the character(s) of the most recently spoken initial phoneme just to the side of a speaker's lips. This design should enable a speechreader to focus on the speaker's eyes and lip movements (as in traditional speechreading), while also monitoring changes in PhonemeViz's state using their peripheral vision to help disambiguate confusing visemes. PhonemeViz can be overlaid onto video or displayed on a transparent head mounted display (Figure 5, left) to augment natural speechreading and enhance speechreading acquisition, although for any real-time applications, new automated speech recognition algorithms tuned to intial phonemes are needed.

In our pilot evaluation of a PhonemeViz prototype [27], PhonemeViz enabled all participants to achieve 100% word recognition (showing successful viseme disambiguation), and participants lauded PhonemeViz in subjective and qualitative feedback. Our initial results demonstrate that visualising a **medium** amount of **analytic** information can improve visual-only speechreading in constrained word recognition tasks.

MirrorTrainer

Our interviews revealed that mirror training plays a key role in speechreading training (used by all seven participants) and is used at home by students. However, traditional mirror training does not fully develop **hybrid** skills as students cannot assess themselves (because they have full knowledge of what they are saying), and the technique trains them to read their own speech (instead of other people). Despite its value in classes, mirror training is not reflected in currently available SATs.

Most mobile devices now have front-facing cameras that could be used to provide speechreaders with a '21st century' mirror. Utilising the front facing camera, the *MirrorTrainer* app would allow speechreaders to practice their own speech movements (with or without additional context) similar to how mirror training is currently used. However, MirrorTrainer could also capture and store videos of speech movements in a 'speech movement library'. Coupled with a simple user input-based 'labelling' function, MirrorTrainer would have a repository of videos containing known speech movement content (Figure 5, middle) that could be used to overcome the 'full knowledge' limitation of current mirror practice. Users could then share their speech movement libraries with each other, overcoming the 'self-training' limitation of traditional mirror training. In addition, MirrorTrainer users could gather coded videos of friends and family, allowing them to practice speechreading on those they speak with most. Such a MirrorTraining app would provide a variable (**low** to **high**) amount of information.

ContextCueView

Scenarios are used by four of our participants, who emphasised that scenarios provide a rich context for speechreading training. However, there are currently no tools to help speechreaders during real-life versions of scenarios used in class. In addition, there are no SATs that provide **high synthetic** information.

In certain situations, phrases and topics can be pre-associated with a given location or situation. These associations can be shown using a constellation diagram [34], in which a text label for the situation is placed in the middle while related topics and phrases radiate out from the situation label. Constella-

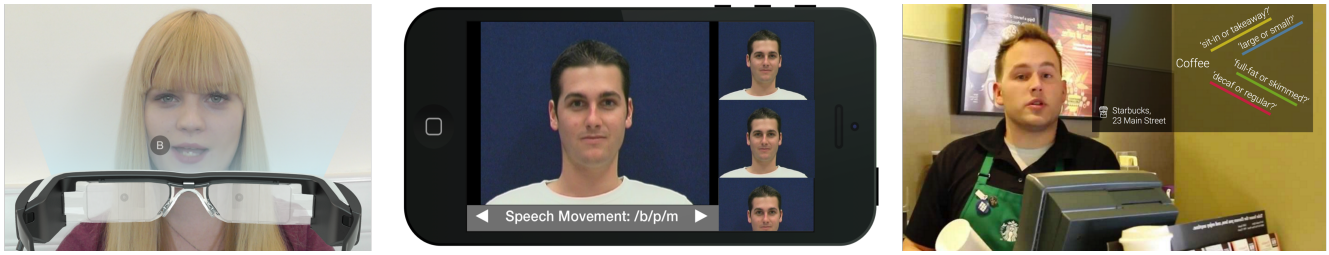

Figure 5. Left to right: PhonemeViz viewed through Epson Moverio glasses (http://www.epson.com/moverio) for 'bat', MirrorTrainer showing multiple words for the speech movement 'p/b/m' (images from vidTIMIT [45]), ContextCueView showing synthetic conversation cues for a coffee shop interaction using Google Glass (https://developers.google.com/glass/).

tion diagrams help by prompting the speechreader to consider potential phrases and topics in advance of a given situation.

ContextCueView would gather contextual data (e.g., GPS, date/time) to anticipate a user's situation. Using this contextual data, ContextCueView would load a matching previously-generated constellation diagram. ContextCueView constellation diagrams would be stored in a central repository, and collectively curated to rapidly provide constellation diagrams for a variety of situations. ContextCueView could run on a mobile device and operate like a 'contextual phrase book', but it is also well-suited for a glanceable display (Figure 5, right).

DISCUSSION

Summary of Contributions

Through thematic analysis of our interviews with speechreading tutors, we identified four main themes relevant to the future development of speechreading: *speechreading as a skill*, limited *access to speechreading*, a broad range of *teaching practices*, and mixed *attitudes to technology*. Using our themes, we developed a novel framework to help design new Speechreading Acquisition Tools (SATs). In evaluating our framework, we demonstrated that it can accommodate current teaching techniques and existing solutions, as well as be used to design new SATs. We developed a prototype of one new SAT named *PhonemeViz* and proposed two additional SATs that we will build and evaluate as future work.

Limitations

First, we based our framework on data obtained exclusively from Scottish speechreading tutors, many of whom were trained on the same course – it is possible that these tutors use outdated techniques, reducing the value of our framework. However, Scotland is one of the few countries in the world that provides accredited speechreading tutor training – most other countries provide no formal training. As such, it is arguable that Scotland's training course is based on best practices.

Second, fitting a technology into our framework does not guarantee that it will be useful for speechreading acquisition. For example, spectrograms provide a high amount of analytic information (which should make them very helpful), but this information is difficult to utilise without considerable training [28]. Likewise, Upton's Eyeglasses and iBaldi provide a low amount of analytic information, but focus on speech production instead of appearance (one of our tutors – a Speech and Language Therapist – indicated that she *"...[switches] off [her] speech therapy brain because [she] would go with sounds but in lipreading it is very much the shape and presentation on the lips."*). Our framework provides substantial guidance for the design of new SATs, but any resulting SATs still need to be evaluated using speechreaders. We argue that this is a strength of our framework not a limitation, as our framework serves to complement and enhance existing best-practice participatory design approaches, not attempt to replace them.

Third, our framework does not distinguish between SATs focussing on speechreading training versus 'live' speechreading. However, as discussed, PhonemeViz and ContextCueView can be extended to 'live' speechreading, suggesting that some SATs can be adapted for both. Therefore, we leave this flexibility within our framework but will consider future refinements to make this distinction more explicit in the future.

Generalisations & Extensions

Our framework currently focusses on English speechreading acquisition, but can be extended to support speechreading in other languages. French [6], German [14], and Japanese [31] each have their own confusing viseme-phoneme mappings, however their speechreading techniques for distinguishing between mouth-shapes (analytic), and conversational repair strategies (synthetic) are similar to English, so our framework should generalise to developing SATs for other languages.

Our framework can also be extended to help develop technology for other skill-based speech domains. For example, speech therapy uses a variety of approaches [20], and already features a number of speech production aids [29, 43]. Likewise, our approach can be extended to language learning, as understanding a foreign language is analytic (e.g., pronunciation) and synthetic (e.g., using context to distinguish homonyms). In particular, ContextCueView (described above) might easily extend to supporting in-situ foreign language conversations.

CONCLUSION & FUTURE WORK

Many people around the world have conversation-impairing hearing loss. Speechreading can dramatically improve conversational understanding, but is a challenging skill to learn. To help design new Speechreading Acquisition Tools (SATs), we developed a framework based on thematic analysis of interviews with speechreading tutors. We then evaluated our framework by using it to classify identified teaching techniques, critique existing solutions, and we demonstrated how to use our framework to design three new SATs. Our future work is focussed on continuing to develop and evaluate our novel SATs in a participatory process involving speechreaders.

REFERENCES

1. Action On Hearing Loss 2011. Facts and figures on deafness and tinnitus. (2011).
2. Action On Hearing Loss 2015. Hearing Matters. https://www.actiononhearingloss.org.uk/supporting-you/policy-research-and-influencing/research/hearing-matters.aspx. (2015).
3. Action On Hearing Loss 2016. Describing Deafness. https://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/deafness/describing-deafness.aspx. (2016).
4. David Armstrong, Ann Gosling, John Weinman, and Theresa Marteau. 1997. The place of inter-rater reliability in qualitative research: an empirical study. *Sociology* 31, 3 (1997), 597–606.
5. L Armstrong. 2015. On everybody's lips. http://www.scotlipreading.org.uk/files/1914/2686/1587/On_everybodys_lips_-_report.pdf. (2015).
6. C. Benoît, M. T. Lallouache, T. Mohamadi, and C. Abry. 1992. A set of French visemes for visual speech synthesis. In *Talking Machines: Theories, Models and Designs*, G. Bailly and C. Benoît (Eds.). Elsevier Science Publishers B. V., North-Holland, Amsterdam, 485–504.
7. H Russell Bernard. 2011. *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira, Lanham, Maryland.
8. CA Binnie. 1977. Attitude changes following speechreading training. *Scandinavian Audiology* 6, 1 (1977), 13–19.
9. Carl A Binnie, Allen A Montgomery, and Pamela L Jackson. 1974. Auditory and visual contributions to the perception of consonants. *Journal of speech, language, and hearing research* 17, 4 (1974), 619–630.
10. Dan G. Blazer, Sarah Domnitz, and Catharyn T. Liverman (Eds.). 2016. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. The National Academies Press, Washington, D.C.
11. Jan Blustein and Barbara E Weinstein. 2016. Opening the Market for Lower Cost Hearing Aids: Regulatory Change Can Improve the Health of Older Americans. *American journal of public health* 106, 6 (2016), 1032–1035.
12. Arthur Boothroyd, Liat Kishon-Rabin, and Robin Waldstein. 1995. Studies of tactile speechreading enhancement in deaf adults. In *Seminars in Hearing*, Vol. 16. Thieme Medical Publishers, Inc., New York, NY, 328–340.
13. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
14. Christoph Bregler and Yochai Konig. 1994. "Eigenlips" for robust speech recognition. In *Acoustics, Speech, and Signal Processing ICASSP-94.*, Vol. 2. IEEE, Piscataway, NJ, II–669.
15. Martha Emma Bruhn. 1949. *The Mueller-Walle Method of Lipreading for The Hard of Hearing*. Volta Bureau, Washington, D.C.
16. Alan Bryman. 2015. *Social research methods*. Oxford University Press, Oxford, United Kingdom.
17. Anna Mae Bunger. 1961. *Speech reading, Jena method: a textbook with lesson plans in full development for hard of hearing adults and discussion of adaptations for hard of hearing and deaf children*. The Interstate, Chicago, IL.
18. Ruth Campbell, Barbara Dodd, and Denis K Burnham. 1998. *Hearing by eye II*. Vol. 2. Psychology Press, Hove, United Kingdom.
19. Richard Orin Cornett. 1967. Cued speech. *Am. Ann. Deaf.* 112, 1 (1967), 3–13.
20. Gary S Dell, Franklin Chang, and Zenzi M Griffin. 1999. Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science* 23, 4 (1999), 517–542.
21. D Ebrahimi and H Kunov. 1991. Peripheral vision lipreading aid. *IEEE transactions on biomedical engineering* 38, 10 (1991), 944–952.
22. Melisa Echalier. 2010. In it together: the impact of hearing loss on personal relationships. (2010).
23. Cletus G Fisher. 1968. Confusions among visually perceived consonants. *Journal of Speech, Language, and Hearing Research* 11, 4 (1968), 796–804.
24. Marilyn French-St George and Richard G Stoker. 1988. Speechreading: An historical perspective. *The Volta Review* 90, 5 (1988), 17–21.
25. Stuart Gatehouse, Graham Naylor, and Clous Elberling. 2003. Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology* 42, sup1 (2003), 77–85.
26. E Goldstein. 2013. *Sensation and perception*. Cengage Learning, Independence, KY.
27. Benjamin M Gorman. 2016. Reducing viseme confusion in speech-reading. *ACM SIGACCESS Accessibility and Computing* 114 (2016), 36–43.
28. Beth G Greene, David B Pisoni, and Thomas D Carrell. 1984. Recognition of speech spectrograms. *JASA* 76, 1 (1984), 32–43.
29. Joshua Hailpern, Karrie Karahalios, and James Halle. 2009. Creating a spoken impact: encouraging vocalization through audio visual feedback in children with ASD. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, ACM, New York, NY, 453–462.
30. Louise Hickson, Carly Meyer, Karen Lovelock, Michelle Lampert, and Asad Khan. 2014. Factors associated with success with hearing aids in older adults. *International journal of audiology* 53, sup1 (2014), S18–S27.

31. Shizuo Hiki and Yumiko Fukuda. 1997. Negative Effect of Homophones on Speechreading in Japanese. In *Audio-Visual Speech Processing: Computational & Cognitive Science Approaches*. ISCA, Baxias, France, 9–12.
32. Janet Jeffers and Margaret Barley. 1980. *Speechreading (lipreading)*. Charles C. Thomas Publisher, Springfield, IL.
33. Carl J Jensema, Ramalinga Sarma Danturthi, and Robert Burch. 2000. Time spent viewing captions on television programs. *Am. Ann. Deaf.* 145, 5 (2000), 464–468.
34. Harriet Kaplan, Scott J Bally, and Carol Garretson. 1985. *Speechreading: A way to improve understanding*. Gallaudet University Press, Chicago, IL.
35. C.E Kinzie and R. Kinzie. 1920. The Kinzie method of speech reading. *Volta Review* 22 (1920), 609–19.
36. Patrick Lucey, Terrence Martin, and Sridha Sridharan. 2004. Confusability of phonemes grouped according to their viseme classes in noisy environments. In *Proc. of Australian Int. Conf. on Speech Science & Tech*. Australasian Speech Science and Technology Association (ASSTA), Canberra City, ACT, Australia, 265–270.
37. David F Marks and Lucy Yardley. 2004. *Research methods for clinical and health psychology*. Sage, Thousand Oaks, CA, United States.
38. Dominic W Massaro, Michael M Cohen, Walter Schwartz, Sam Vanderhyden, and Heidi Meyer. 2013. Facilitating Speech Understanding for Hearing-Challenged Perceivers in Face-to-Face Conversation and Spoken Presentations. *ICTHP* (2013).
39. Laura Matthews. 2011. Unlimited potential?: a research report into hearing loss in the workplace. (2011).
40. Abby McCormack and Heather Fortnum. 2013. Why do people fitted with hearing aids not wear them? *International Journal of Audiology* 52, 5 (2013), 360–368.
41. Edward Bartlett Nitchie. 1919. *Lip-reading Principles and Practise: A Hand-book for Teachers and for Self Instruction*. Frederick A. Stokes Company, New York, NY.
42. Catherine Pope, Sue Ziebland, and Nicholas Mays. 2000. Analysing qualitative data. *Bmj* 320, 7227 (2000), 114–116.
43. Zoe Roxburgh, James M Scobbie, and Joanne Cleland. 2015. Articulation therapy for children with cleft palate using visual articulatory models and ultrasound biofeedback. *Proceedings of the 18th ICPhS, Glasgow* 0858 (2015).
44. Oliver Sacks. 2009. *Seeing voices: A journey into the world of the deaf*. Pan Macmillan, London, United Kingdom.
45. Conrad Sanderson and Brian C Lovell. 2009. Multi-region probabilistic histograms for robust and scalable identity inference. In *International Conference on Biometrics*. Springer, New York, NY, 199–208.
46. June E Shoup. 1980. Phonological aspects of speech recognition. In *Trends in speech recognition*. Prentice Hall, Englewood Cliffs, NJ, 125–138.
47. Donald G. Sims, C Dorn, C Clark, L Bryant, and B Mumford. 2002. New developments in computer assisted speechreading and auditory training. *Paper presented at the American Speech-Language Hearing Association convention* (2002).
48. Nancy Tye-Murray. 2002. Conversation made easy: Speechreading and conversation training for individuals who have hearing loss (adults and teenagers). *St. Louis: Central Institute for the Deaf* (2002).
49. Nancy Tye-Murray. 2014. *Foundations of aural rehabilitation: Children, adults, and their family members*. Nelson Education, Scarborough, ON, Canada.
50. University of Auckland, School of Psychology 2016. Why don't we advocate multiple-coders and inter-rater reliability for TA? https://www.psych.auckland.ac.nz/en/about/our-research/research-groups/thematic-analysis/frequently-asked-questions-8.html#c83c77d6d1c625135085e489bd66e765. (2016).
51. Hubert W Upton. 1968. Wearable eyeglass speechreading aid. *American Annals of the Deaf* 113, 2 (1968), 222–229.
52. Akira Watanabe, Shingo Tomishige, and Masahiro Nakatake. 2000. Speech visualization by integrating features for the hearing impaired. *IEEE Trans. Speech Audio Process.* 8, 4 (2000), 454–466.
53. WHO 2015. Deafness and hearing loss, Fact sheet N.300. http://www.who.int/mediacentre/factsheets/fs300/en/. (2015).
54. Lei Xie, Yi Wang, and Zhi-Qiang Liu. 2006. Lip Assistant: Visualize Speech for Hearing Impaired People in Multimedia Services. In *Proc. SMC'06*, Vol. 5. IEEE, Piscataway, NJ, 4331–4336.
55. Hanfeng Yuan, Charlotte M Reed, and Nathaniel I Durlach. 2005. Tactual display of consonant voicing as a supplement to lipreading. *The Journal of the Acoustical society of America* 118, 2 (2005), 1003–1015.